

## **Adopt a web-enabled data abstraction standard such as w10n “webification”**

Chase Million (Million Concepts LLC, chase.million@gmail.com)

The w10n “webification” protocol was developed by JPL researchers for the web-based abstraction of data and metadata formats. In essence, it allows end users to mine data structures as well as the files themselves (including, for example, header data and, in the case of images, individual pixel values). The Imaging Node has already implemented w10n for their holdings [1], and the technology is at the heart of their Marsviewer tool. We have found it to be incredibly useful and powerful. I suggest that the PDS adopt it more widely.

After learning about the adoption of the w10n implementation by IMG at the 2015 Data Workshop in Flagstaff [2], I started using this service for an ongoing research project working with Pancam data. It has *significantly* improved our work in a number of ways. We are able to mine the header data in an automated, fast, and repeatable way without being required to do a bulk download of the data to store locally. Further, while a new project to support PDS3 image formats in Python was born of that very workshop, and looks to be promising, at the time of the workshop we had actually built our own bare bones tool to read these files. Webification allows the image data to be returned as easily parseable JSON data structures, eliminating our need to parse the data format at all. Due to local storage and network limitations, we could only perform analysis on small sets of the data at a time, and, because our work involved mining the data for very rare events, we could not test our methods on truly representative subsets of the data, and we feared that we would not be able to scale the analysis to the full data set. But webification solves this problem because it allows us to quickly and automatically pull down the data precisely when and as it was needed.

The w10n protocol has also dramatically simplified the workflow of our remote collaboration because we did not need to either pass large image files back and forth to have discussions about the same data, nor did we need to independently locate and download those image files locally. We could simply send code in an email that sent a w10n request for what we wanted and perform all of the analysis right in the language. This made iteration on the work---software debugging to data analysis---much, much more time and resource efficient than it would otherwise have been.

Creating a tool or library to read out a format in one language does not necessarily translate well to other languages. As language preferences of the research community evolve, and as formats or standards in PDS change over time, new tools have to be developed or old tools need to be adopted, and that process takes both time and money. For example, I’m aware of recent issues the PDS has had with supporting the CDF format, and a consequent request in the most recent PDART call for standardized CDF libraries. While I absolutely applaud the shift in PDS4 to XML label files, one researcher at the 2015 Data Workshop presented on the challenges of creating an appropriate XML parser for IDL. I have experience with astronomers unable to read PDS3 .IMG formatted files and planetary scientists unable to read the common-to-astronomy .FITS file format (which is also archived in the PDS), even though both formats are extremely similar “under the hood.”

These problems can effectively vanish, for the vast majority of end users, with the adoption of w10n across the PDS. The great beauty of w10n “webification” standard is that any

programming language of choice for the researcher only requires the ability to (1) send and receive HTTP requests and (2) parse JSON formatted data. These are two extremely, extremely common required functionalities for modern software development and, as such, are implemented almost immediately (and robustly) in new languages.

The adoption of w10n across the PDS would not only reduce time to value for locating and working with PDS data, but it would also dramatically increase the number of tools that could easily work with that data (as, e.g., different programming languages), potentially increasing the overall scientific return.

[1] <http://data.jpl.nasa.gov/planetary-help>

[2] Xing, Z., and E. Sayfi. "Webification (W10N)-Data on the Web Platform." *LPI Contributions* 1846 (2015): 7066.